



HƯỚNG DẪN ĐỌC TOÀN VĂN BÁO CÁO KQNC



😊 **BẠN MUỐN ĐỌC NHANH
NHỮNG THÔNG TIN CẦN THIẾT ?**

☞ **Hãy đọc qua Mục lục bên tay trái bạn trước khi
đọc báo cáo** (với Acrobat 4.0 trở lên, cho trỏ chuột vào
mỗi đề mục để đọc toàn bộ dòng bị che khuất)

☞ **Chọn đề mục muốn đọc và nháy chuột vào đó**

😊 **BẠN MUỐN PHÓNG TO HAY THU NHỎ
TRANG BÁO CÁO TRÊN MÀN HÌNH ?**

☞ **Chọn, nháy chuột vào 1 trong 3 kích thước
có sẵn trên thanh Menu**



, hoặc

☞ **Mở View trên thanh Menu, Chọn Zoom to**

☞ **Chọn tỷ lệ có sẵn trong hộp kích thước
hoặc tự điền tỷ lệ theo ý muốn, Nhấn OK**

*Chúc bạn hài lòng
với những thông tin được cung cấp*

Bộ phận quản trị CSDL toàn văn KQNC
P.508, 24 Lý Thường Kiệt, Hà Nội
Tel.(04)-9-349-126

R

**ỦY BAN CHỨNG KHOÁN NHÀ NƯỚC
TRUNG TÂM NGHIÊN CỨU KHOA HỌC
VÀ BỒI DƯỠNG NGHIỆP VỤ CHỨNG KHOÁN**

ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP BỘ NĂM 2000

MÃ SỐ :

**XÂY DỰNG VÀ PHÁT TRIỂN
CƠ SỞ DỮ LIỆU
CHỨNG KHOÁN VIỆT NAM**

**Chủ nhiệm đề tài : TS. Tôn tích Quý
Phó giám đốc trung tâm**

Hà nội, tháng 12 năm 2000

3795^A

1416/01

Lời giới thiệu

Trung tâm giao dịch chứng khoán đầu tiên ở nước ta đã được khai trương ngày 20/7/2000. Đây là kết quả của sự nỗ lực, cố gắng của toàn ngành chứng khoán Việt nam và nó không thể tách rời khỏi sự quan tâm chỉ đạo của Đảng, Chính phủ, sự ủng hộ nhiệt tình của các Bộ, ban ngành, các địa phương và sự hưởng ứng, đón tiếp nhiệt thành của các doanh nghiệp, cũng như đông đảo công chúng đầu tư.

Thông tin thị trường được ví như mạch máu, nguồn năng lượng nuôi sống thị trường. Do đó, để tạo điều kiện cho thị trường hoạt động tốt và phát triển, chúng ta không thể không chú ý đến việc xây dựng một hệ thống thông tin chứng khoán đầy đủ và nhanh nhạy, nhằm phục vụ tốt nhất cho công tác điều hành, quản lý thị trường, phục vụ nhu cầu thông tin cho công chúng đầu tư, cho việc nghiên cứu, học tập và đào tạo phát triển nguồn nhân lực cho thị trường chứng khoán Việt nam. Trong chiến lược thông tin tổng thể đó thì việc xây dựng và phát triển một cơ sở dữ liệu (CSDL) chứng khoán hoàn chỉnh về mặt nội dung, linh hoạt trong tra cứu, hợp lý theo trình tự phát triển của thị trường đang là vấn đề đang nổi lên đặc biệt cần quan tâm, chú ý.

Tất cả các nước có thị trường chứng khoán đều xây dựng cho riêng mình cơ sở dữ liệu. Tất cả các cơ sở dữ liệu đó nhìn chung đã được điện tử hoá và phần lớn được hoà chung vào mạng thông tin toàn cầu và chúng ta có thể truy cập và nghiên cứu, tìm hiểu thông tin một cách dễ dàng. Đối với Việt nam cũng vậy, chúng ta cũng nên thiết kế xây dựng trang Web chứng khoán này. Tuy nhiên không phải mọi thông tin đều được công khai hết. Có thông tin chỉ phải bảo mật trong một thời gian nhất định, có những thông tin lại bí mật lâu dài hơn, có những thông tin chỉ có một nhóm người được biết mà thôi. Nói như vậy có nghĩa là mỗi loại thông tin " mật " đều có nhu cầu bảo mật riêng mà trong quá trình thiết kế, xây dựng cơ sở dữ liệu luôn luôn cần phải chú ý, quan tâm.

Đối với nước ta, thị trường chứng khoán mới ra đời, việc đưa đề tài này vào nghiên cứu trong năm nay đã có muộn tý chút, nhưng nhìn chung là phù hợp với tiến độ phát triển chung của thị trường.

Việc xây dựng cơ sở dữ liệu chứng khoán Việt nam được thực hiện sau nhiều nước khác trên thế giới và sau khi Việt nam đã xây dựng thành công một số cơ sở dữ liệu khác là một thuận lợi lớn. Chúng

ta vừa có thể tiếp thu những kinh nghiệm của các cơ sở dữ liệu chứng khoán của các nước khác trên thế giới lại vừa tiếp thu được kinh nghiệm xây dựng các cơ sở dữ liệu khác của Việt nam. Đặc biệt là đề tài đã mời được chuyên gia đầu ngành về cơ sở dữ liệu ở nước ta tham gia thực hiện, người đã trực tiếp tham gia, chỉ huy xây dựng nhiều cơ sở dữ liệu khác ở Việt nam, đó là PGS. TS Nguyễn Đức Thi - Trưởng phòng cơ sở dữ liệu - Viện công nghệ thông tin - Trung tâm khoa học xã hội nhân văn quốc gia tham gia. Đó là những đảm bảo để cơ sở dữ liệu này vừa đảm bảo tính đầy đủ, tính khoa học hệ thống và tính nhanh nhạy trong thực tiễn, vừa đảm bảo tính tối ưu trong tra cứu, tìm kiếm. Trong số các cơ sở dữ liệu khác đã được xây dựng ở nước ta, chúng ta không thể không kể đến :

- + CSDL quốc gia về tài chính, ngân sách thuộc Bộ tài chính
- + CSDL quốc gia về kinh tế XH thuộc Tổng cục Thống kê
- + CSDL về dân cư thuộc UB dân số và kế hoạch hoá gia đình
- + CSDL về luật pháp thuộc Bộ tư pháp
- + CSDL về đất đai và tài nguyên thuộc Tổng cục địa chính
- + CSDL về công chức và đối tượng hưởng chính sách thuộc Ban tổ chức và cán bộ Chính phủ

Đề tài này nhằm các mục đích chính là thực hiện nghiên cứu tiền khả thi cho CSDL CK VN : Trong đó nghiên cứu cả mặt lý luận, thực tiễn, giải pháp và các điều kiện để thực hiện.

Nội dung của đề tài gồm các nhóm vấn đề chính như sau :

- Một số phần lý luận có tính chất giới thiệu về lý thuyết CSDL
- Tìm hiểu kinh nghiệm xây dựng hệ thống thông tin thị trường chứng khoán của một số nước trên thế giới.
- Tìm hiểu kinh nghiệm xây dựng một số CSDL lớn của nước ta.
- Hệ thống hoá hệ thống thông tin của thị trường chứng khoán Việt nam, Dự báo khối lượng thông tin TTCK qua các thời kỳ phát triển của TTCK Việt nam;
- Thiết kế hệ thống cho CSDL chứng khoán Vn, chú ý các mặt : Các yêu cầu khi xây dựng; Các đơn vị tham gia; 4 modul : Modul dữ liệu; Modul trao đổi dữ liệu và modul xử lý dữ liệu và modul quản trị CSDL
- Đề xuất giải pháp phần mềm, hạ tầng kỹ thuật cho CSDL CK VN
- Đề xuất dự kiến nhu cầu nguồn nhân lực cho việc vận hành, duy trì và cập nhật CSDL CK VN

Đề tài tập trung chính vào việc phân tích hệ thống thông tin thị trường chứng khoán Việt nam và đề ra giải pháp về cấu trúc hệ thống, cùng các biện pháp thực hiện.

Đề tài này là đề tài rộng, lớn, rất nhiều nội dung. Bởi vậy để tránh cho đề tài quá dài chúng tôi đã sử dụng các công cụ để rút gọn nó lại như sử dụng dấu ..vv, chữ viết tắt mà mọi người có thể quen biết và nhận ra ngay, như CSDL (Cơ sở dữ liệu), TTCK (thị trường chứng khoán), TTGDCK TP. Hồ Chí Minh (Trung tâm giao dịch chứng khoán Thành phố Hồ Chí Minh), HTTT (hệ thống thông tin), CNTT (công nghệ thông tin).

Chúng tôi xin chân thành cảm ơn sự góp ý chỉ đạo của Hội đồng xét duyệt đề cương do Chủ tịch Ủy ban Chứng khoán Nhà nước , TS. Nguyễn Đức Quang làm Chủ tịch Hội đồng. Cảm ơn sự cộng tác đặc lực, rất hiệu quả của PGS. TS Nguyễn Đức Thi, Trưởng phòng cơ sở dữ liệu (CSDL), Viện công nghệ thông tin, Trung tâm khoa học tự nhiên và công nghệ quốc gia, chuyên gia đầu ngành về cơ sở dữ liệu của Việt nam. Cảm ơn sự cộng tác của Giám đốc – TS Đào Lê Minh và sự cộng tác của các cộng tác viên khác ở trong và ngoài Trung tâm. Chúng tôi tin chắc rằng, nếu thiếu sự cộng tác đó thì đề tài này không thể hoàn thành được.

Hà nội tháng 11 năm 2000

CHƯƠNG I

CƠ SỞ LÝ LUẬN, THỰC TIỄN XÂY DỰNG CSDL CHỨNG KHOÁN VIỆT NAM

I- Giới thiệu về cơ sở lý luận của cơ sở dữ liệu

1- Một số vấn đề cơ bản về CSDL

a) Khái niệm về cơ sở dữ liệu và công dụng.

Hiểu một cách mô tả, cơ sở dữ liệu là tập hợp hệ thống thông tin, tư liệu được sắp xếp ngăn nắp, khoa học trong một hệ thống nhằm tạo điều kiện thuận lợi nhất cho việc cập nhật, bổ sung, khai thác, tra cứu, kết xuất và tổng hợp thông tin. Ngày nay hầu như các cơ sở dữ liệu đều đã được tin học hoá nhờ thành quả to lớn của bộ môn khoa học công nghệ thông tin.

Trong thời kỳ đầu để tạo điều kiện cho việc tự động hoá xử lý dữ liệu người ta đã sử dụng các tệp dữ liệu, nó là nơi chứa thông tin và người ta dùng các chương trình để tìm kiếm, thao tác trên các dữ liệu của tệp đó. Đó là tiền thân, tiền đề ban đầu của các hệ thống cơ sở dữ liệu. Để xử lý dữ liệu một cách khoa học cũng đòi hỏi phải sử dụng cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu thường có phương pháp, công cụ để lưu trữ, tìm kiếm, sửa đổi và chuyển đổi các dữ liệu. Đó là các chức năng đầu tiên, được thực hiện theo các thuật toán hoàn thiện, đảm bảo dữ liệu được bảo vệ và đảm bảo việc thao tác chiết xuất, xử lý đạt hiệu quả cao.

Mô tả dữ liệu được xem như việc xác định *tên, dạng dữ liệu và tính chất* của dữ liệu. Hệ quản trị cơ sở dữ liệu khác hẳn với hệ quản lý các tệp hay các tệp cơ sở dữ liệu bởi lẽ nó cho phép mô tả dữ liệu theo cách không phụ thuộc vào người sử dụng, không phụ thuộc vào yêu cầu tìm kiếm và thay đổi thông tin.

Việc tổ chức một hệ thống thông tin hay xây dựng một cơ sở dữ liệu cho các ngành khoa học, ngành kinh tế đã trở nên phổ biến từ những năm 90. Điều đó chứng tỏ khả năng ứng dụng rộng rãi của ngành khoa học này. Nhiều cơ sở dữ liệu ngành hay quốc gia đã được thiết lập. Các ứng dụng tiêu biểu của cơ sở dữ liệu có thể kể ra như:

- Tổ chức thông tin trong các bài toán khoa học kỹ thuật

- Kho dữ liệu trong các hệ thống thông tin quản lý
- Tổ chức dữ liệu có cấu trúc phức tạp như các dữ liệu địa lý
- Cơ sở dữ liệu trong các hệ thống hỗ trợ công nghiệp, hỗ trợ giảng dạy
- Tổ chức thông tin đa phương tiện, xử lý tri thức.

Thứ nhất- Cơ sở dữ liệu ứng dụng trong các bài toán khoa học kỹ thuật. Các bài toán này có thuật toán khó, thường thì không đòi hỏi công cụ tốt nhất về tổ chức dữ liệu. Tuy nhiên, trong các bài toán phức tạp hơn, với nhiều dữ liệu trung gian thì cách tổ chức dữ liệu hợp lý là điều không thể không nghĩ đến.

Thứ hai- Ứng dụng của cơ sở dữ liệu trong quản lý. Công tác quản lý không cần thuật toán phức tạp, nhưng thường đòi hỏi xử lý nhiều dữ liệu. Khối lượng lớn thông tin cần được tổ chức khoa học để tiện cho quá trình xử lý.

Thứ ba- Ứng dụng thông tin trong các ngành khoa học không phải là công nghệ thông tin, thí dụ như vật lý, hoá học, sinh học, ngôn ngữ,... ở các lĩnh vực này cũng có các nhu cầu lưu trữ, xử lý dữ liệu. Các cơ sở dữ liệu riêng biệt này mang những đặc tính riêng của từng ngành. Các dữ liệu về địa lý, bao gồm các bảng số, các ảnh, các phương pháp truy nhập đến các kho dữ liệu ... rất cần được tổ chức và xử lý hợp lý. Các dữ liệu địa lý, địa chất, thủy văn, môi trường,... thường đòi hỏi các phương tiện nhớ có dung lượng lớn và được xử lý trên các bộ xử lý đặc biệt để đảm bảo tốc độ cao.

Thứ tư- Cơ sở dữ liệu ứng dụng trong hệ thống hỗ trợ. Việc tổ chức và xử lý dữ liệu cũng có nhu cầu trong các ứng dụng có sử dụng hệ chuyên gia, người máy, xử lý các quá trình công nghiệp. Như vậy, cơ sở dữ liệu cần có khả năng cơ giới hoá việc tìm kiếm thông tin nhờ cơ chế suy luận tự động. Vấn đề thời gian thực trong cơ sở dữ liệu được giải quyết để phù hợp với các hệ thống công nghiệp. Thời gian có thể được thể hiện trong cơ sở dữ liệu thông qua hai cách:

- Thời gian tương đối trong hệ quản trị cơ sở dữ liệu, liên quan đến thay đổi trạng thái của cơ sở dữ liệu
- Thời gian tuyệt đối của môi trường được mô tả trong cơ sở dữ liệu, liên quan đến trạng thái của môi trường

Khoa học về cơ sở dữ liệu còn dùng để tổ chức cơ sở tri thức, thiết lập hệ thống câu hỏi, chọn mô hình trong hệ thống hỗ trợ giảng dạy, hay trong công nghệ dạy học.

Cuối cùng- cơ sở dữ liệu ứng dụng trong hệ thống đa phương tiện, xử lý tri thức: *Giao diện người dùng* trong cơ sở dữ liệu, đề cập các nghiên cứu về quan hệ và sự kiện, đề cập việc tổ chức các câu hỏi cho người sử dụng. Người ta nhận thấy không có ngôn ngữ nào là đặc biệt quan trọng và ưu điểm vượt trội, ngay cả ngôn ngữ đồ thị. Một giao diện hiển thị thường được ưu chuộng, với khả năng:

- Đưa ra câu trả lời cho các câu hỏi dưới dạng hiển thị như đồ thị, lược đồ, có tác dụng nhấn mạnh trực giác
- "Lật trang" của một cơ sở dữ liệu, tức là khả năng lựa chọn thông tin nhanh một cách tự nhiên
- Tìm kiếm trong cơ sở dữ liệu theo phương thức con người đã quen thuộc, chẳng hạn như cách tìm sách trong các tủ sách của thư viện

Trong số các giao diện người dùng, giao diện đa hình thái (multimodal), giao diện dùng ngôn ngữ tự nhiên được quan tâm và nay cũng có nhiều kết quả đáng khích lệ.

Tuy không được xây dựng như hệ thống tri thức hay hệ chuyên gia, cơ sở dữ liệu có thể mô tả và xử lý các tri thức. Một thế hệ mới của các cơ sở dữ liệu đang được xây dựng với các hệ thống cơ sở dữ liệu đa dạng, trong đó có cơ sở dữ liệu suy diễn. Các tri thức xử lý được thể hiện dưới dạng:

- Tri thức tổng quát như các luật và sự kiện
- Các điều kiện thay đổi, hoặc kích hoạt dữ liệu
- Suy diễn các thông tin có liên hệ với các sự kiện và luật

Ngoài ra, người ta còn đề cập khía cạnh về xử lý các tri thức không đầy đủ.

Thông thường một CSDL thường có một số mô hình sau (Modul)

a) Mô hình dữ liệu

Thông thường đối với việc thiết kế và xây dựng các hệ thống thông tin quản lý, chúng ta cần xây dựng các tệp dữ liệu. Các tệp này bao gồm nhiều bản ghi có cùng một cấu trúc xác định (loại bản ghi). Mỗi bản ghi được phân thành các trường dữ liệu. Như vậy, *một cơ sở dữ liệu là một hệ thống các tệp dữ liệu. Mỗi tệp này có cấu trúc bản ghi khác nhau nhưng về mặt nội dung lại có quan hệ với nhau. Lúc đó, một hệ quản trị cơ sở dữ liệu là một hệ thống quản lý và điều hành các tệp dữ liệu*. Nói chung một hệ quản trị cơ sở dữ liệu thường có những đặc tính sau:

- Có tính độc lập với các công cụ lưu trữ.
- Có tính độc lập với các chương trình phần mềm của người sử dụng (có nghĩa là các ngôn ngữ lập trình khác nhau có thể được dùng trong hệ này),
- Có khả năng tại một thời điểm truy nhập vào nhiều nơi trong hệ
- Có khả năng khai thác tốt tiềm năng của máy
- Người dùng có kiến thức tối thiểu cũng dùng được
- Bảo đảm an toàn dữ liệu và bảo mật dữ liệu
- Thuận lợi và mềm dẻo trong việc bổ sung, loại bỏ, thay đổi dữ liệu

-Giảm bớt sự dư thừa dữ liệu trong lưu trữ

Trong quá trình thiết kế và xây dựng các hệ cơ sở dữ liệu, người ta tiến hành xây dựng các mô hình dữ liệu. Mô hình dữ liệu phải thể hiện được các mối quan hệ bản chất của các dữ liệu mà các dữ liệu này phản ánh các mối quan hệ và các thực thể trong thế giới thực. Có thể thấy mô hình dữ liệu phản ánh khía cạnh cấu trúc logic mà không đi vào khía cạnh vật lý của các cơ sở dữ liệu. Khi xây dựng các mô hình dữ liệu cần phân biệt các thành phần cơ bản sau:

-*Thực thể*: Đó là đối tượng có trong thực tế mà chúng ta cần mô tả các đặc trưng của nó

-*Thuộc tính*: Đó là các dữ liệu thể hiện các đặc trưng của thực thể

-*Ràng buộc*: Đó là mối quan hệ logic của các thực thể

Tuy vậy, ba thành phần cơ bản trên được thể hiện ở hai mức:

-*Mức loại dữ liệu (type)*: Đó là sự khái quát hoá các ràng buộc, các thuộc tính, các thực thể cụ thể

-*Mức thể hiện*: Đó là một ràng buộc cụ thể, hoặc là các giá trị thuộc tính, hoặc là một thực thể cụ thể

Thông thường chúng ta nhận được các loại dữ liệu của các đối tượng cần khảo sát trong quá trình phân tích và các thể hiện cụ thể của chúng.

Yếu tố quan trọng nhất của cấu trúc cơ sở dữ liệu là **dạng** cấu trúc dữ liệu mà trong đó các mối quan hệ giữa các dữ liệu lưu trữ được mô tả. Có thể thấy rằng loại dữ liệu nền tảng của việc mô tả các mối quan hệ là **loại bản ghi (Record type)**. Bởi vì các ràng buộc giữa các loại bản ghi tạo ra bản chất cấu trúc của cơ sở dữ liệu. Vì thế dựa trên việc xác định các ràng buộc giữa các loại dữ liệu được cho như thế nào mà chúng ta phân loại các mô hình dữ liệu. Có nghĩa là từ cách nhìn của người sử dụng, việc mô tả các dữ liệu và các ràng buộc giữa các dữ liệu được thực hiện như thế nào để phân biệt các cơ sở dữ liệu. Trên thực tế người ta phân biệt 3 loại mô hình dữ liệu :

-*Mô hình phân cấp*: dữ liệu được phân loại theo những tiêu chuẩn khác nhau và được cấu trúc theo kiểu phân cấp dạng cây. Cứ mỗi tiêu chuẩn phân loại xác định một tầng trong hệ thống phân cấp.

-*Mô hình dữ liệu mạng*: thể hiện trực tiếp các ràng buộc tùy ý giữa các loại bản ghi

-*Mô hình dữ liệu quan hệ*: các ràng buộc được thể hiện qua các quan hệ (bảng)

Mô hình dữ liệu quan hệ là một công cụ rất tiện lợi để mô tả cấu trúc logic của các cơ sở dữ liệu. Như vậy, ở mức logic mô hình này bao gồm các tệp được biểu diễn dưới dạng các bảng. Do đó đơn vị của cơ

sở dữ liệu quan hệ là một bảng, trong đó các dòng của bảng là các bản ghi dữ liệu cụ thể (các thể hiện cụ thể của loại bản ghi) còn tên các cột là các thuộc tính

Theo cách nhìn của người sử dụng thì một cơ sở dữ liệu quan hệ là một tập hợp các bảng biến đổi theo thời gian.

Trong tiến trình nghiên cứu cơ sở dữ liệu thì ngoài việc nghiên cứu chính bản thân cơ sở dữ liệu, chúng ta còn cần phải nghiên cứu hệ quản trị cơ sở dữ liệu cùng với ngôn ngữ thể hiện trong cơ sở dữ liệu đó để xử lý dữ liệu .

Trong quá trình thiết kế hệ thống thông tin, việc xây dựng các mô hình dữ liệu theo ngôn ngữ "thực thể -quan hệ" là rất quan trọng để đảm bảo tính thống nhất và khả năng trao đổi, tìm kiếm thông tin trong các CSDL

Trên cơ sở các mô hình dữ liệu, cần xác định hệ thống các CSDL phục vụ cho các phân hệ nêu trên. Hệ thống này cần được tổ chức và quản lý một cách thống nhất để dùng chung cho toàn hoạt động điều hành và quản lý.

Các CSDL được tập trung trong các Server và có thể truy nhập được từ mọi máy trạm.

Mô hình dữ liệu sẽ :

-Mô tả các dữ liệu chính liên quan đến các đối tượng, các tệp cần được quản lý trong hệ thống. Mỗi đối tượng, văn bản có thể thuộc về một chủ đề phân loại nào đó

-Mô tả các dữ liệu-được lưu giữ trong hệ thống

b) Mô hình trao đổi thông tin

Các nội dung chính của mô hình bao gồm:

- Mô tả đường đi của thông tin vào, ra
- Mô tả nội dung thông tin vào ra, cập nhật sửa đổi, bổ sung, copy
- Mô tả thời gian trao đổi
- Mô tả phân hệ: Xét xem phân hệ thông tin hỗ trợ về thông tin nhằm phục vụ có hiệu quả lĩnh vực quản lý nào, Các chức năng chính của phân hệ này là những gì.

- Mô tả các đối tượng: Cấp trên, lãnh đạo Bộ, các đơn vị trong Ủy ban, các đơn vị trực thuộc, các đơn vị địa phương.

- Mô tả các chức năng nghiệp vụ: quản lý những cái gì. Mỗi chức năng bao gồm các hoạt động nghiệp vụ nào.

- Mô tả các thông báo: các loại thông tin như : Hồ sơ, bổ sung, báo cáo dự trù, chính sách, đơn thư, yêu cầu thông tin, đề nghị, tổng kết, kế hoạch,...